

ABSTRACT OF THE DISCLOSURE

The browsable database can allow for high-throughput analysis of protein sequences. One helpful feature may be a simplified ontology of protein function, which allows browsing of the database by biological functions. Biologist curators may have associated the ontology terms with Hidden Markov Models (HMMs), rather than individual sequences, so that they can be applied to additional sequences. To ensure accurate functional classification, HMMs may be constructed not only for families, but for curator-defined subfamilies, whenever family members have divergent functions or nomenclature. Multiple sequence alignments and phylogenetic trees, including curator-assigned information, can be available for each family. Various versions of the browsable database may include training sequences from all organisms in the GenBank non-redundant protein database, and the HMMs can be used to classify gene products across the entire genomes of human, and *Drosophila melanogaster*.